

12.2. Locating heavy-atom sites

BY M. T. STUBBS AND R. HUBER

12.2.1. The origin of the phase problem

Once a native data set has been collected, the next task is the solution of the structure. There is one major hurdle: the phase problem. To study objects at the atomic level, we must utilize waves with a wavelength in the ångström range, *i.e.* X-radiation. X-rays interact with electrons and so provide an image of the electron distribution of the sample. Unfortunately, X-rays are refracted by matter only very weakly, and so it is not possible to construct a lens to view molecules at atomic dimensions.*

As shown in Chapter 2.1, the diffraction $F(\mathbf{S})$ obtained from an electron-density distribution $\rho(\mathbf{r})$ is given by

$$F(\mathbf{S}) = \int \rho(\mathbf{r}) \exp\{2\pi i \mathbf{r} \cdot \mathbf{S}\} d^3 \mathbf{r},$$

where \mathbf{S} is perpendicular to the scattered wave and $|\mathbf{S}| = 2 \sin \theta / \lambda$; θ is the scattering angle and λ is the wavelength. The diffraction pattern is a *Fourier transform* of the electron density. If we have a crystal with cell parameters a , b and c , then the Laue diffraction conditions require that \mathbf{S} lies on a reciprocal lattice such that $\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$, where \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* are the reciprocal-lattice vectors, and h , k and l are the integer indices of the diffracted beam.

$$F(hkl) = V \sum_{xyz} \rho(xyz) \exp\{2\pi i(hx + ky + lz)\},$$

where V represents the volume of the unit cell, and x , y and z are the fractional coordinates within that cell in the directions of \mathbf{a} , \mathbf{b} and \mathbf{c} .

Since the diffraction pattern is a Fourier transform of the electron density, it follows that the electron density is an inverse Fourier transform of the diffraction pattern:

$$\rho(\mathbf{r}) = \int F(\mathbf{S}) \exp\{-2\pi i \mathbf{r} \cdot \mathbf{S}\} d^3 \mathbf{S},$$

$$\rho(xyz) = (1/V) \sum_{hkl} F(hkl) \exp\{-2\pi i(hx + ky + lz)\}.$$

Thus it should be mathematically straightforward to calculate the electron density from the diffraction pattern. This is, unfortunately, not the case. The function $F(\mathbf{S})$ describing the diffracted rays is a complex function with a *magnitude* $|F(\mathbf{S})|$ and a *phase* $\varphi(\mathbf{S})$. The diffraction experiment measures the *intensities* $I(\mathbf{S})$, however; the relationship between $I(\mathbf{S})$ and $F(\mathbf{S})$ is:

$$I(\mathbf{S}) = F^*(\mathbf{S}) \cdot F(\mathbf{S}) = |F(\mathbf{S})|^2,$$

where $F^*(\mathbf{S})$ is the complex conjugate of $F(\mathbf{S})$. The measured intensities are related directly to the magnitudes of the diffracted beams; the phase information, however, is lost (Fig. 12.2.1.1): this is the origin of the *phase problem*.

There are essentially four ways of overcoming the phase problem (Fig. 12.2.1.2):

(1) the use of isomorphous replacement to influence the diffraction pattern, thereby revealing information about the phases (this chapter);

(2) Patterson search techniques, which in essence allow modelling of the phase distribution of an unknown crystal based upon that of a known molecule (Part 13);

(3) the use of anomalous dispersion, which shifts the imaginary component of $F(\mathbf{S})$, allowing an experimental measurement of the phase (Part 14); and

(4) direct methods, which make use of probabilistic relationships between different diffracted rays (Part 16).

The method of isomorphous replacement, by which the first macromolecular structures were solved (Green *et al.*, 1954), remains the most widely used technique for *ab initio* structure determination, although the availability of synchrotrons, with their facility for selecting a desired wavelength, and molecular-biology techniques that allow the direct introduction of anomalous scatterers, such as selenium or tellurium, into the protein of interest (Hendrickson *et al.*, 1990; Budisa *et al.*, 1997) have proven that multiple anomalous dispersion is an exceptionally powerful technique for the solution of novel structures. Patterson search techniques (Rossmann, 1972) are ideal if a similar macromolecular structure is already known, while direct methods are more-or-less confined to very high resolution data (Sheldrick, 1990).

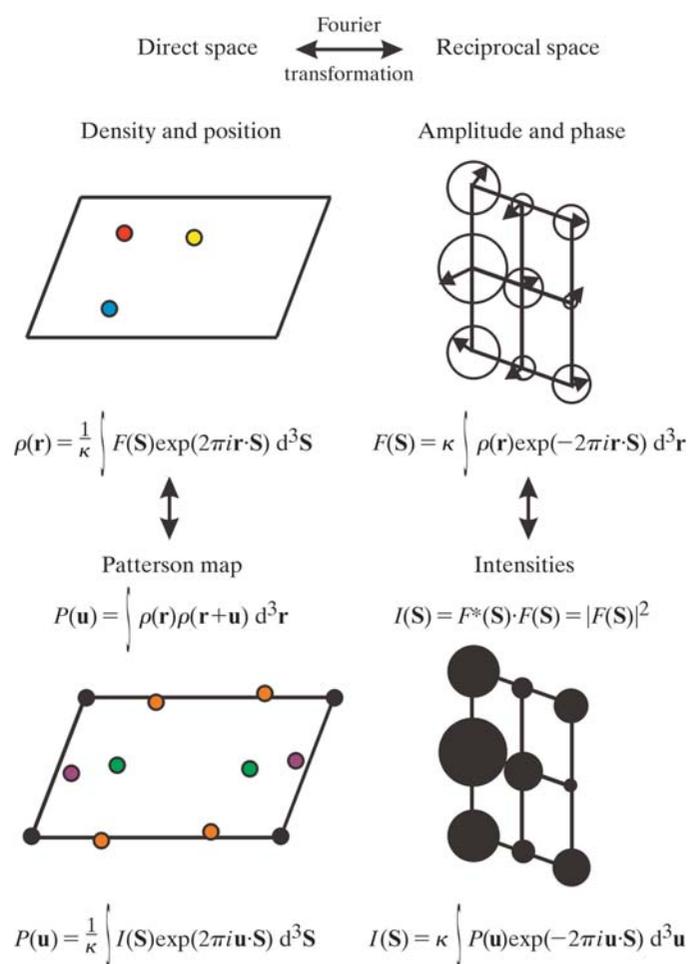


Fig. 12.2.1.1. Relationships in diffraction space. The diffraction pattern of an object is the Fourier transform of the electron density, consisting of both amplitudes and phases. What we measure in an X-ray diffraction experiment, however, are the diffracted intensities; the phase information is lost. The Fourier transform of the intensities results in the Patterson map, which is related to the electron density as follows. For any two atoms in the structure, the vector between them, centred at the origin, has a value corresponding to the product of their densities. Thus, the red atom and the yellow atom result in the orange cross vectors, the red and blue atoms result in the magenta cross vectors, and the yellow and blue atoms result in green cross vectors. As each atom has a 'cross vector' to itself, a large peak is found at the origin; the Patterson map is centrosymmetric.

* This is not strictly true, due to advances in electron microscopy (see Part 19); however, this method has yet to find the universality enjoyed by X-ray crystallography.

12.2. LOCATING HEAVY-ATOM SITES

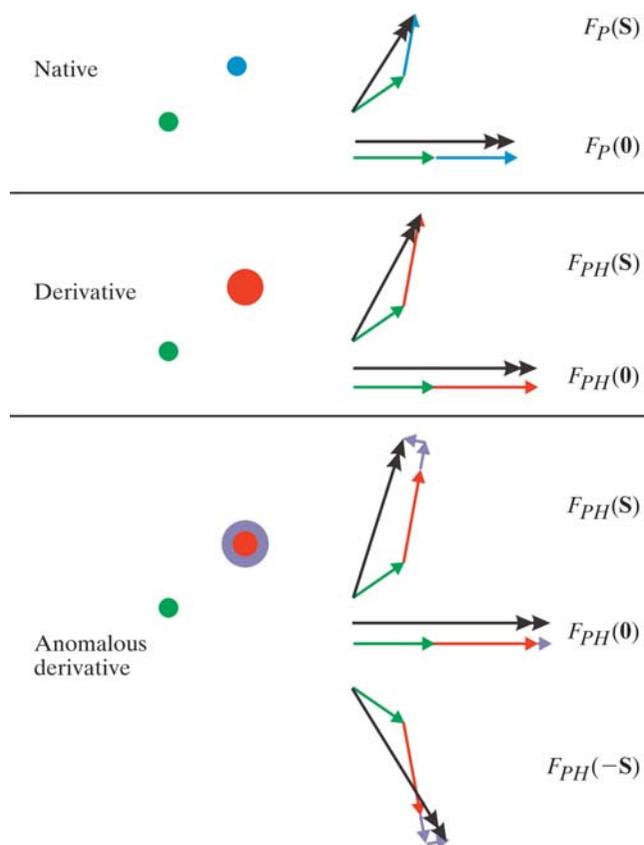


Fig. 12.2.1.2. The effect of introducing a heavy atom or anomalous scatterer. The native two-atom structure gives rise to two diffraction vectors (green and blue) of equal magnitude but different phase (see Chapter 2.1), with a resultant diffraction vector F_P (black). Isomorphous replacement of the blue atom by the larger red one gives rise to a diffraction vector of greater magnitude but equivalent phase (red), causing a change in the resultant magnitude F_{PH} (and hence the intensity) and in the phase. Introduction of an anomalous scatterer results in a phase shift (lilac) of the diffraction vector, resulting in differing amplitudes and phases for $F_{PH}(S)$ and $F_{PH}(-S)$.

In order to obtain phase information from isomorphous replacement (or from anomalous dispersion), it is necessary to locate the atomic positions of the heavy-atom (or anomalous) scatterers.

12.2.2. The Patterson function

Although the set of measured intensities contains no information regarding the phases, the Fourier transform of the intensities, the so-called Patterson function, contains valuable information. Patterson (1934) showed that the inverse Fourier transform of the intensity,

$$P(uvw) = (1/V) \sum_{hkl} I(hkl) \exp\{-2\pi i(hu + kv + lw)\},$$

is related to the electron density by

$$P(\mathbf{u}) = \int \rho(\mathbf{r})\rho(\mathbf{r} + \mathbf{u}) d^3\mathbf{r}.$$

The Patterson function $P(\mathbf{u})$ is an autocorrelation function of the density. For every vector \mathbf{u} that corresponds to an interatomic vector, $P(\mathbf{u})$ will contain a peak (Fig. 12.2.1.1). These are some properties of the Patterson function:

(1) Every atom makes an 'interatomic vector' with itself, and therefore the *origin peak*, $P(\mathbf{0}) = \sum \rho^2(\mathbf{r})$, dominates the Patterson function. This origin peak can be 'removed' through subtraction of the average intensity from $I(hkl)$ before Fourier transformation.

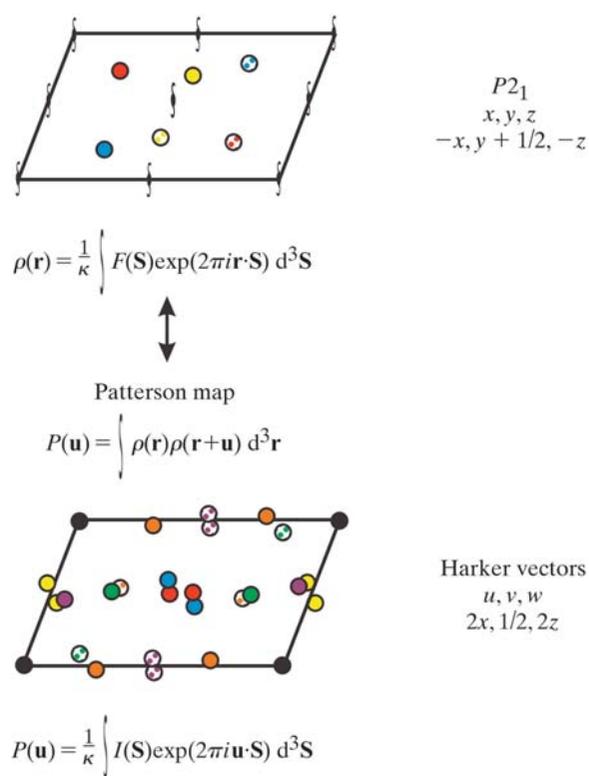


Fig. 12.2.2.1. The Patterson map with symmetry. When the crystal unit cell contains more than one molecule, then additional cross vectors will be formed between differing molecules. If these are related by crystallographic symmetry, there is a geometrical relationship between cross peaks. In this diagram, the peaks of Fig. 12.2.1.1 are supplemented by those between atoms of symmetry-related molecules. The red, yellow and blue peaks of the resulting Patterson function represent those between same atoms (*i.e.* red to red, yellow to yellow and blue to blue) related by symmetry. These peaks are found on a Harker section.

(2) For every vector between $\rho_i(\mathbf{r}_i)$ and $\rho_j(\mathbf{r}_j)$, the same value (*i.e.* their product) is found for $\rho_j(\mathbf{r}_j)$ to $\rho_i(\mathbf{r}_i)$, and so the Patterson map is *centrosymmetric*.

(3) For a structure consisting of n atoms, there are $n(n-1)/2$ cross vectors, and so the Patterson function is extremely crowded.

For simple crystals, the Patterson map can be used to solve the structure directly. For macromolecular structures, the Patterson map provides a vehicle for solving the phase problem.

If the crystal contains rotational symmetry elements, then the cross vectors between $\rho_i(\mathbf{r}_i)$ and its symmetry mate lie on a plane perpendicular to the symmetry axis – the *Harker section* (Harker, 1956). By way of example, the space group $P2_1$ has two symmetry-related positions (Fig. 12.2.2.1),

$$(x, y, z) \text{ and } (-x, y + \frac{1}{2}, -z).$$

Cross vectors between symmetry-related points will therefore have the form

$$(2x, \frac{1}{2}, 2z),$$

i.e. all cross vectors lie on the plane $v = \frac{1}{2}$. For space group $P2_12_12_1$, the general coordinates

$$(x, y, z), (x + \frac{1}{2}, -y + \frac{1}{2}, -z), (-x + \frac{1}{2}, -y, z + \frac{1}{2}), (-x, y + \frac{1}{2}, -z + \frac{1}{2})$$

give rise to cross vectors

$$(\frac{1}{2}, 2y + \frac{1}{2}, 2z), (2x + \frac{1}{2}, 2y, \frac{1}{2}), (2x, \frac{1}{2}, 2z + \frac{1}{2}),$$

12. ISOMORPHOUS REPLACEMENT

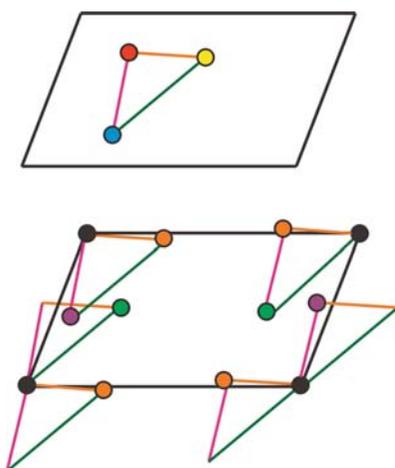


Fig. 12.2.2.2. The vector superposition method. The Patterson map of Fig. 12.2.1.1 can be regarded as the superposition of the structure (and its inverse), with each of its atoms placed alternately at the origin. By shifting each peak of the Patterson function to the origin and calculating the correlation of all remaining peaks with the unshifted map, it is possible to deconvolute the Patterson function.

i.e. there are three Harker sections: $u = \frac{1}{2}$, $v = \frac{1}{2}$ and $w = \frac{1}{2}$. Peaks occurring on the Harker sections must reduce to a self-consistent set of coordinates (x, y, z) , allowing reconstruction of the atomic positions.

If we have two isomorphous (see below) data sets F_{PH} and F_P , then the difference in the two Patterson functions,

$$P_{PH} - P_P = \int [F_{PH}^2(\mathbf{S}) - F_P^2(\mathbf{S})] \exp\{-2\pi i \mathbf{r} \cdot \mathbf{S}\} d^3\mathbf{S},$$

will deliver information about the heavy-atom structure. Such a difference function gives rise to non-negligible peaks arising from interference between the F_H and F_P terms, however (Perutz, 1956). Rossmann (1960) showed that these interference terms could be reduced through calculation of the modified Patterson function

$$P_H = \int [F_{PH}(\mathbf{S}) - F_P(\mathbf{S})]^2 \exp\{-2\pi i \mathbf{r} \cdot \mathbf{S}\} d^3\mathbf{S}.$$

In the case of a single-site derivative, peaks should occur only at the Harker vectors corresponding to the heavy-atom position. Even so, there is a choice of positions for the heavy atom: *e.g.*, in the $P2_12_12_1$ case, coordinates $(\pm x + \xi, \pm y + \nu, \pm z + \zeta)$, where ξ , ν and ζ can each take the value 0 or $1/2$, will all give rise to the same Harker vectors. This in itself is not a problem, relating to equivalent choices of origin and of handedness, but has important ramifications for multisite derivatives or multiple isomorphous replacement (see below).

If there is more than one site, then there will be two sets of peaks: one set corresponding to the Harker sections (self-vector set) and one set corresponding to the difference vectors between different heavy-atom sites (the cross-vector set). In this case, the choice of one heavy-atom position (x_{H1}, y_{H1}, z_{H1}) determines the origin and the handedness to which all other peaks *must* correspond. Thus, in the $P2_12_12_1$ example, only one cross vector will occur for

$$(x_{h1} \pm x_{h2} + \xi, y_{h1} \pm y_{h2} + \nu, z_{h1} \pm z_{h2} + \zeta).$$

An alternative to the Harker-vector approach is Patterson-vector superposition (Sheldrick *et al.*, 1993; Richardson & Jacobson, 1987). The Patterson map contains several images of the structure that have been shifted by interatomic vectors (Fig. 12.2.2.2). If this structure is relatively simple (as is to be hoped for in a 'normal' heavy-atom derivative), then it should be possible to deconvolute the superimposed structures by vector shifts (Buerger, 1959).

12.2.3. The difference Fourier

Once the heavy-atom positions have been found, they can be used to calculate approximate phases and Fourier maps. Ideally, difference Fourier maps calculated with phases from a single site should reveal the other positions determined from the Harker search procedure. This ensures that all heavy-atom positions correspond to a single origin and hand. Similarly, phases calculated from derivative $H1$ should reveal the heavy-atom structure for derivative $H2$. Merging and refinement of all phase information will result in a phase set that can be used to solve the structure.

12.2.4. Reality

12.2.4.1. Treatment of errors

Until now, we have dealt with cases involving perfect data. Although this ideal may now be attainable using MAD techniques, this is not necessarily the usual laboratory situation. In the first place, it is necessary to scale the derivative data F_{PH} to the native F_P . One of the most common scaling procedures is based on the expected statistical dependence of intensity on resolution (Wilson, 1949). This may not be particularly accurate when only low-resolution data are available, in which case a scaling through equating the Patterson origin peaks of native and derivative sets may provide better results (Rogers, 1965).

A model to account for errors in the data, determination of heavy-atom positions *etc.* was proposed by Blow & Crick (1959), in which all errors are associated with $|F_{PH}|_{\text{obs}}$ (Fig. 12.2.4.1); a more detailed treatment has been provided by Terwilliger & Eisenberg

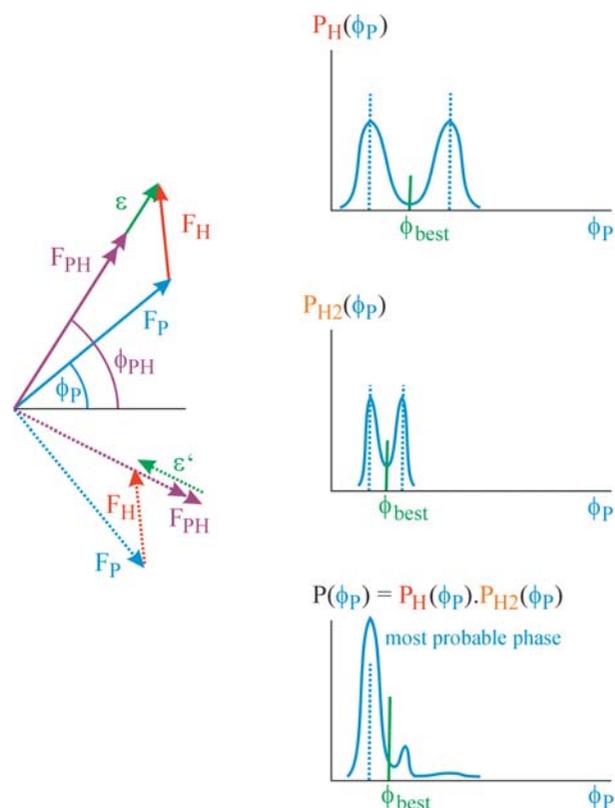


Fig. 12.2.4.1. The treatment of phase errors. The calculated heavy-atom structure results in a calculated value for both the phase and magnitude of F_H (red). According to the value of φ_P , the triangle $F_P - F_H - F_{PH}$ will fail to close by an amount ϵ , the lack of closure (green). This gives rise to a phase distribution which is bimodal for a single derivative. The combined probability from a series of derivatives has a most probable phase (the maximum) and a best phase (the centroid of the distribution), for which the overall phase error is minimum.

12.2. LOCATING HEAVY-ATOM SITES

(1987). Owing to errors, the triangle formed by F_P , F_{PH} and F_H fails to close. The *lack of closure error* ε is a function of the calculated phase angle φ_P :

$$\varepsilon(\varphi_P) = |F_{PH}|_{\text{obs}} - |F_{PH}|_{\text{calc}}.$$

Once an initial set of heavy-atom positions has been found, it is necessary to refine their parameters (x , y , z , occupancy and thermal parameters). This can be achieved through the minimization of

$$\sum_S \varepsilon^2/E,$$

where E is the estimated error ($\simeq \langle (|F_{PH}|_{\text{obs}} - |F_{PH}|_{\text{calc}})^2 \rangle$) (Rossmann, 1960; Terwilliger & Eisenberg, 1983). This procedure is safest for noncentrosymmetric reflections (φ restricted to 0 or π) if enough are present. Phase refinement is generally monitored by three factors:

$$R_{\text{Cullis}} = \sum ||F_{PH} + F_P| - |F_H|_{\text{calc}}| / \sum |F_{PH} - F_P|$$

for noncentrosymmetric reflections only; acceptable values are between 0.4 and 0.6;

$$R_{\text{Kraut}} = \sum ||F_{PH}|_{\text{obs}} - |F_{PH}|_{\text{calc}}| / \sum |F_{PH}|_{\text{obs}},$$

which is useful for monitoring convergence; and the

$$\text{phasing power} = \sum |F_H|_{\text{calc}} / \sum ||F_{PH}|_{\text{obs}} - |F_{PH}|_{\text{calc}}|,$$

which should be greater than 1 (if less than 1, then the phase triangle cannot be closed *via* F_H).

The resulting phase probability is given by

$$P(\varphi_P) = \exp\{-\varepsilon^2(\varphi_P)/2E^2\}.$$

The phases have a minimum error when the *best phase* φ_{best} , *i.e.* the centroid of the phase distribution,

$$\varphi_{\text{best}} = \int \varphi_P P(\varphi_P) d\varphi_P,$$

is used instead of the most probable phase. The quality of the phases is indicated by the *figure of merit* m , where

$$m = \int P(\varphi_P) \exp(i\varphi_P) d\varphi_P / \int P(\varphi_P) d\varphi_P.$$

A value of 1 for m indicates no phase error, a value of 0.5 represents a phase error of about 60° , while a value of 0 means that all phases are equally probable.

The *best Fourier* is calculated from

$$\rho_{\text{best}}(\mathbf{r}) = (1/V) \sum_{\mathbf{S}} m |F_P(\mathbf{S})| \exp\{i\varphi_{P\text{best}}(\mathbf{S})\},$$

where the electron density should have minimal errors.

12.2.4.2. Automated search procedures

If the derivative shows a high degree of substitution, then the Harker sections become more difficult to interpret. Furthermore, Terwilliger *et al.* (1987) have shown that the intrinsic noise in the difference Patterson map increases with increasing heavy-atom substitution. It is at this stage that automated procedures are invaluable.

One such automated procedure is implemented in *PROTEIN* (Steigemann, 1991). The unit cell is scanned for possible heavy-atom sites; for each search point (x , y , z), all possible Harker vectors are calculated, and the difference-Patterson-map values at these points are summed or multiplied. As the origin peak dominates the Patterson function, this region is set to zero. The resulting correlation map should contain peaks at all possible heavy-atom positions. The peak list can then be used to find a set of consistent heavy-atom locations through a subsequent search for difference vectors (cross vectors) between putative sites. It should be possible

to locate all major and minor heavy-atom sites through repetition of this procedure. A similar strategy is adopted in the program *HEAVY* (Terwilliger *et al.*, 1987), but sets of heavy-atom sites are ranked according to the probability that the peaks are not random. The program *SOLVE* (Terwilliger & Berendzen, 1999) takes this process a stage further, where potential heavy-atom structures are solved and refined to generate an (interpretable) electron density in an automated fashion.

The search method can also be applied in reciprocal space, where the Fourier transform of the trial heavy-atom structure is calculated, and the resulting $F_{H\text{calc}}$ is compared to the measured differences between derivative and native structure-factor amplitudes (Rossmann *et al.*, 1986). In the programme *XtalView* (McRee, 1998), the correlation coefficient between $|F_H|$ and $|F_{PH} - F_P|$ is calculated, whilst a correlation between F_H^2 and $(F_{PH} - F_P)^2$ is used by Badger & Athay (1998). Dumas (1994b) calculates the correlation between $|F_{H\text{calc}}|^2$ and $|F_{H\text{estimated}}|^2$, based on the estimated lack of isomorphism.

Vagin & Teplyakov (1998) have reported a heavy-atom search based on a reciprocal-space translation function. In this case, low-resolution peaks are not removed but weighted down using a Gaussian function. Potential solutions are ranked not only according to their translation-function height, but also through their phasing power, which appears to be a stronger selection criterion.

All these searches are based upon the sequential identification of heavy-atom sites and their incorporation in a heavy-atom partial structure. Problems arise when bogus sites influence the search for further heavy-atom positions. In an attempt to overcome this problem, the heavy-atom search has been reprogrammed using a genetic algorithm, with the Patterson minimum function as a selection criterion (Chang & Lewis, 1994). This approach has the potential to reveal all heavy-atom positions in one calculation, and tests on model data have shown it to be faster than traditional sequential searches.

12.2.5. Special complications

12.2.5.1. Lack of isomorphism

This problem is by far the most common in protein crystallography. An isomorphous derivative is one in which the crystalline arrangement has not been disturbed by derivatization. An early study of Crick & Magdoff (1956) proposed a rule of thumb that a change in any of the cell dimensions by more than around 5% would result in a lack of isomorphism that would defeat any attempt to locate the heavy-atom positions or extract useful phase information. Lack of isomorphism can, however, be more subtle; sometimes a natural variation in the native crystal form may occur, resulting in poor merging statistics of data obtained from different crystals. Coupling this variation with commonly observed structural changes upon heavy-atom binding can provide a considerable barrier to obtaining satisfactory phases. Dumas (1994a) has provided a theoretical consideration of this problem.

One practical approach is to collect native and derivative data sets from the same crystal, a technique that has been successful in the structure determination of cyclohydrolase (Nar *et al.*, 1995), proteosome (Löwe *et al.*, 1995) and a number of other proteins. Nonisomorphism can be used, however. In the structure solution of carbamoyl sarcosine hydrolase (Romao *et al.*, 1992), derivatives fell into two (related) crystalline classes. By judicious use of two 'native' crystal forms, heavy-atom positions could be obtained in each of the two classes. Phasing and resultant averaging between the two classes provided an interpretable electron density. In the case of ascorbate oxidase (Messerschmidt *et al.*, 1989), multiple isomorphous replacement failed to provide an interpretable density. It was possible, however, to place the initial density into a second

12. ISOMORPHOUS REPLACEMENT

crystal form, which in turn provided phases of sufficient quality to determine heavy-atom sites in derivatives of the second form. Phase-combination and density-modification techniques in the two crystal forms allowed the solution of the structure.

12.2.5.2. Space-group problems

Although the macromolecular crystallographer is rarely confronted with the problems facing their small-molecule colleagues with regard to determining the correct space group, the simplified heavy-atom structure may often throw some surprises. Certain pseudosymmetries may become 'exact' for the heavy-atom difference Patterson map. Thus, cross peaks between different heavy atoms may occur on a Harker section (or 'pseudo-Harker section'), complicating interpretation of the Patterson map. Such was the case with azurin (Adman *et al.*, 1978; Nar *et al.*, 1991), where the heavy-atom structure gave rise to a pseudo-homometric Patterson function, *i.e.* one in which two possible (nonequivalent) choices were available for the heavy-atom structure, only one of which was correct. This arose from a pseudo-centring of the lattice that became almost exact for the heavy-atom structure.

In the case of human NC1 (Stubbs *et al.*, 1990), all heavy-atom derivatives appeared to lie on or near the crystallographic twofold axis. This resulted in a partially centrosymmetric heavy-atom structure that failed to deliver sufficient phase information for noncentrosymmetric reflections. To check for problems with the native data set, anomalous difference Patterson maps {coefficients $[F_{PH}(\mathbf{S}) - F_{PH}(-\mathbf{S})]^2$ } were calculated. Coincidence of the peaks obtained from conventional and anomalous Patterson syntheses showed that the heavy-atom positions were correct, but unfortunately did not lead to a structure solution.

12.2.5.3. High levels of substitution; noncrystallographic symmetry

Most problematic are the cases where many heavy atoms have become incorporated in the asymmetric unit. Not only does this

cause difficulties in the scaling of derivative to native data, but also the large number of peaks results in ambiguities in the solution of the Patterson function. In such cases, it may be necessary to obtain primary phase information from a different source (such as, for example, another low-substitution-site derivative). One important subclass of high-level substitution is when the native asymmetric unit contains several copies of a single molecule (noncrystallographic symmetry or NCS).

A major problem in locating complex noncrystallographic axes is that the geometrical relationship between NCS peaks in the Patterson map is nontrivial. Under certain conditions, NCS results in a recognizable local symmetry within the Patterson map (Stubbs *et al.*, 1996). In many cases, however, these conditions (that the NCS axes of crystallographic symmetry-related molecules are parallel) are not fulfilled. Under such circumstances, all heavy-atom sites (including all crystallographic symmetry-related positions) must be checked carefully with the rotation function in order to pinpoint the NCS axis. This is relatively trivial for low-order NCS (twofold, threefold), but becomes increasingly complicated for higher orders. It should also always be borne in mind that the heavy-atom positions might not necessarily follow the NCS constraints due to crystal packing. If there is reason to suspect that sites are related by local symmetry, then the orientation of this axis can be used in the initial Harker searches; in practice, however, such searches are extremely sensitive to the correct orientation of the axis.

In the case of high-order NCS (such as, *e.g.*, with icosahedral virus structures or symmetric macromolecular complexes), an alternative approach to the usual initial Harker-vector search can be provided by the self-rotation function. Knowledge of the orientation of the NCS axis (from the rotation function) can be used to determine the relative positions of heavy atoms to the NCS axis (Argos & Rossmann, 1976; Arnold *et al.*, 1987; Tong & Rossmann, 1993). The orientation can be refined and the resulting peaks can be used as input in a subsequent translation search of the Harker sections.

References

12.1

- Abraham, D. J., Phillips, S. E. V. & Kennedy, P. E. (1983). *Methylphenylmercury: a novel heavy atom reagent for protein crystallography*. *J. Mol. Biol.* **170**, 249–252.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures*. *J. Mol. Biol.* **112**, 534–552.
- Blake, C. C. F. (1968). *The preparation of isomorphous derivatives*. In *Advances in protein chemistry*, Vol. 23, pp. 59–120. New York and London: Academic Press.
- Blundell, T. L. (1968). Unpublished results.
- Blundell, T. L. & Jenkins, J. A. (1977). *The binding of heavy metals to proteins*. *Chem. Soc. Rev. (London)*, **6**, 139–171.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein crystallography*. New York: Academic Press.
- Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neufeind, T., Moroder, L. & Huber, R. (1997). *Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins*. *J. Mol. Biol.* **271**, 1–8.
- Carvin, D. G. A. (1986). Unpublished results.
- Carvin, D. G. A., Islam, S. A., Sternberg, M. J. E. & Blundell, T. L. (1991). *Isomorphous replacement and anomalous scattering*. Warrington: Daresbury Laboratory.
- Ceska, T. A., Sayers, J. R., Stier, G. & Suck, D. (1996). *A helical arch allowing single-stranded DNA to thread through t5 5'-exonuclease*. *Nature (London)*, **382**, 90–93.
- Chen, W. & Bahl, O. P. (1991). *Recombinant carbohydrate and selenomethionyl variants of human choriogonadotropin*. *J. Biol. Chem.* **266**, 8192–8197.
- Doublé, S. (1997). *Preparation of selenomethionyl proteins for phase determination*. *Methods Enzymol.* **276**, 523–530.
- Ely, K. R., Girling, R. L., Schiffer, M., Cunningham, D. E. & Edmundson, A. B. (1973). *Preparation and properties of crystals of a Bence-Jones dimer with mercury inserted into the interchain disulphide bond*. *Biochemistry*, **12**, 4233–4237.
- Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. E. (1994). *Biological Macromolecular Crystallization Database, version 3.0: new features, data and the NASA archive for protein crystal growth data*. *Acta Cryst.* **D50**, 408–413.
- Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The method of isomorphous replacement for protein crystallography*. *Proc. R. Soc. London Ser. A*, **225**, 287–307.
- Hendrickson, W. A. (1985). *Analysis of protein structure from diffraction measurement at multiple wavelengths*. *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

REFERENCES

12.1 (cont.)

- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672.
- Islam, S. A., Carvin, D., Sternberg, M. J. E. & Blundell, T. L. (1998). HAD, a data bank of heavy-atom binding sites in protein crystals: a resource for use in multiple isomorphous replacement and anomalous scattering. *Acta Cryst. D* **54**, 1199–1206.
- Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, O. L. & Ellenberger, T. E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell*, **86**, 321–329.
- Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P. & Hendrickson, W. A. (1994). Crystallization of a fragment of human fibronectin: introduction of methionine by site-directed mutagenesis to allow phasing via selenomethionine. *Proteins Struct. Funct. Genet.* **19**, 48–54.
- Lipka, J. J., Lippard, S. J. & Wall, J. S. (1976). Visualisation of polymercurimethane-labelled fd bacteriophage in the scanning transmission electron microscope. *Science*, **206**, 1419–1421.
- Lipscomb, W. N., Reeke, G. N., Hartsuck, J. A., Quijcho, F. A. & Bethge, P. H. (1970). The structure of carboxypeptidase a. VIII. Atomic interpretation at 0.2 nm resolution, a new study of the complex of glycyl-L-tyrosine with CPA, and mechanistic deductions. *Philos. Trans. R. Soc. London Ser. B*, **257**, 177–214.
- Lustbader, J. W., Wu, H., Birken, S., Pollak, S., Kolks-Gawinowicz, M. A., Pound, A. M., Austen, D., Hendrickson, W. A. & Canfield, R. E. (1995). The expression, characterization and crystallization of wild-type and selenomethionyl human chorionic gonadotrophin. *Endocrinology*, **136**, 640–650.
- Miller, J. H. (1972). Cysteine auxotrophic strains, in which proteins can be synthesised with the seleno derivative. In *Experiments in molecular genetics*. Cold Spring Harbour Laboratory Press.
- Muller, S., Senn, H., Gsell, B., Vetter, W., Baron, C. & Bock, A. (1994). The formation of diselenide bridges in proteins by incorporation of selenocysteine residues: biosynthesis and characterization of (Se)2-thioredoxin. *Biochemistry*, **33**, 3404–3412.
- O'Halloran, T. V., Lippard, S. J., Richmond, T. J. & Klug, A. (1987). Multiple heavy-atom reagents for macromolecular X-ray structure determination application to the nucleosome core particle. *J. Mol. Biol.* **194**, 705–712.
- Petsko, G. A., Phillips, D. C., Williams, R. J. P. & Wilson, I. A. (1978). On the protein crystal chemistry of chloroplatinite ions: general principles and interactions with triose phosphate isomerase. *J. Mol. Biol.* **120**, 345–359.
- Ringe, D., Petsko, G. A., Yamakura, F., Suzuki, K. & Ohmori, D. (1983). Structure of iron superoxide dismutase from *Pseudomonas ovalis* at 2.9 Å resolution. *Proc. Natl Acad. Sci. USA*, **80**, 3879–3883.
- Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. & Blundell, T. L. (1975). X-ray analysis of glucagon and its relationship to receptor binding. *Nature (London)*, **257**, 751–757.
- Schiltz, M. (1997). Xenon & krypton at LURE. <http://www.lure.u-psud.fr/sections/Xenon/XENONENG.HTM>.
- Schoenborn, B. P., Watson, H. C. & Kendrew, J. C. (1965). Binding of xenon to sperm whale myoglobin. *Nature (London)*, **207**, 28–30.
- Sigler, P. B. & Blow, D. M. (1965). A means of promoting heavy atom binding in protein crystals. *J. Mol. Biol.* **14**, 640–644.
- Sigler, P. B., Blow, D. M., Matthews, B. W. & Henderson, R. (1968). Structure of crystalline alpha-chymotrypsin II. A preliminary report including a hypothesis for the activation mechanism. *J. Mol. Biol.* **35**, 143–164.
- Sperling, R., Burstein, Y. & Steinberg, I. Z. (1969). Selective reduction and mercuration of cysteine IV–V in bovine pancreatic ribonuclease. *Biochemistry*, **8**, 3810–3820.
- Sperling, R. & Steinberg, I. Z. (1974). Simultaneous reduction and mercuration of disulphide bond A6–A11 of insulin by monovalent mercury. *Biochemistry*, **13**, 2007–2013.
- Thygesen, J., Weinstein, S., Franceschi, F. & Yonath, A. (1996). The suitability of metal clusters for phasing in macromolecular crystallography of large macromolecular assemblies. *Structure*, **4**, 513–518.
- Tsernoglou, D. & Petsko, G.-A. (1976). The crystal structure of a post-synaptic neurotoxin from sea snake at 2.2 Å resolution. *FEBS Lett.* **68**, 1–4.
- Wood, S. P., Tickle, I. J., Treharne, A. M., Pitts, J. E., Mascarenhas, Y., Li, J. Y., Husain, J., Cooper, S., Blundell, T. L., Hruby, V. J., Buku, A., Fischman, A. J. & Wyssbrod, H. R. (1986). Crystal structure analysis of deamino-oxytocin: conformational flexibility and receptor binding. *Science*, **232**, 633–636.
- Yonath, A., Saper, M. A., Makowski, I., Mussig, J., Piefke, J., Bartunik, H. D., Bartels, K. S. & Wittmann, H. G. (1986). Characterization of single crystals of the large ribosomal particles from *Bacillus stearothermophilus*. *J. Mol. Biol.* **187**, 633–636.

12.2

- Adman, E. T., Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1978). A crystallographic model for azurin at 3.0 Å resolution. *J. Mol. Biol.* **123**, 35–47.
- Argos, P. & Rossmann, M. G. (1976). A method to determine heavy-atom positions for virus structures. *Acta Cryst. B* **32**, 2975–2983.
- Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. & Rossmann, M. G. (1987). The structure determination of a common cold virus, human rhinovirus 14. *Acta Cryst. A* **43**, 346–361.
- Badger, J. & Athay, R. (1998). Automated and graphical methods for locating heavy-atom sites for isomorphous replacement and multiwavelength anomalous diffraction phase determination. *J. Appl. Cryst.* **31**, 270–274.
- Blow, D. M. & Crick, F. H. C. (1959). The treatment of errors in the isomorphous replacement method. *Acta Cryst.* **12**, 794–802.
- Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neufeind, T., Moroder, L. & Huber, R. (1997). Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins. *J. Mol. Biol.* **270**, 616–623.
- Buerger, M. J. (1959). *Vector space*. New York: Wiley.
- Chang, G. & Lewis, M. (1994). Using genetic algorithms for solving heavy-atom sites. *Acta Cryst. D* **50**, 667–674.
- Crick, F. H. C. & Magdoff, B. S. (1956). The theory of the method of isomorphous replacement for protein crystals. I. *Acta Cryst.* **9**, 901–908.
- Dumas, P. (1994a). The heavy-atom problem: a statistical analysis. I. A priori determination of best scaling, level of substitution, lack of isomorphism and phasing power. *Acta Cryst. A* **50**, 526–537.
- Dumas, P. (1994b). The heavy-atom problem: a statistical analysis. II. Consequences of the a priori knowledge of the noise and heavy-atom powers and use of a correlation function for heavy-atom-site determination. *Acta Cryst. A* **50**, 537–546; erratum, **A50**, 793.
- Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc. R. Soc. London Ser. A*, **225**, 287–307.
- Harker, D. (1956). The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Cryst.* **9**, 1–9.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672.
- Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science*, **268**, 533–539.
- McRee, D. E. (1998). *Practical protein crystallography*. San Diego: Academic Press.

12. ISOMORPHOUS REPLACEMENT

12.2 (cont.)

- Messerschmidt, A., Rossi, A., Ladenstein, R., Huber, R., Bolognesi, M., Gatti, G., Marchesini, A., Petruzzelli, R. & Finazzi-Agro, A. (1989). X-ray crystal structure of the blue oxidase ascorbate oxidase from zucchini. Analysis of the polypeptide fold and a model of the copper sites and ligands. *J. Mol. Biol.* **206**, 513–529.
- Nar, H., Huber, R., Meining, W., Schmid, C., Weinkauff, S. & Bacher, A. (1995). Atomic structure of GTP cyclohydrolase I. *Structure*, **3**, 459–466.
- Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M. & Canters, G. W. (1991). X-ray crystal structure of the two site-specific mutants His35Gln and His235Leu of azurin from *Pseudomonas aeruginosa*. *J. Mol. Biol.* **218**, 427–447.
- Patterson, A. L. (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.* **46**, 372–376.
- Perutz, M. F. (1956). Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Cryst.* **9**, 867–873.
- Richardson, J. W. & Jacobson, R. A. (1987). In *Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi. Oxford University Press.
- Rogers, D. (1965). In *Computing methods in crystallography*, edited by J. S. Rollett, pp.133–148. Oxford University Press.
- Romao, M. J., Turk, D., Gomis-Ruth, F. X., Huber, R., Schumacher, G., Mollering, H. & Russmann, L. (1992). Crystal structure analysis, refinement and enzymatic reaction mechanism of *N*-carbamoylsarcosine amidohydrolase from *Arthrobacter* sp. at 2.0 Å resolution. *J. Mol. Biol.* **226**, 1111–1130.
- Rossmann, M. G. (1960). The accurate determination of the position and shape of heavy-atom replacement groups in proteins. *Acta Cryst.* **13**, 221–226.
- Rossmann, M. G. (1972). Editor. *The molecular replacement method*. New York: Gordon and Breach.
- Rossmann, M. G., Arnold, E. & Vriend, G. (1986). Comparison of vector search and feedback methods for finding heavy-atom sites in isomorphous derivatives. *Acta Cryst.* **A42**, 325–334.
- Sheldrick, G. M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. *Acta Cryst.* **A46**, 467–473.
- Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. (1993). The application of direct methods and Patterson interpretation to high-resolution native protein data. *Acta Cryst.* **D49**, 18–23.
- Steigemann, W. (1991). Recent advances in the PROTEIN program system for the X-ray structure analysis of biological macromolecules. In *Crystallographic computing 5: from chemistry to biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 115–125. Oxford University Press.
- Stubbs, M. T., Nar, H., Löwe, J., Huber, R., Ladenstein, R., Spangfort, M. D. & Svensson, L. A. (1996). Locating a local symmetry axis from Patterson map cross vectors: application to crystal data from GroEL, GTP cyclohydrolase I and the proteosome. *Acta Cryst.* **D52**, 447–452.
- Stubbs, M. T., Summers, L., Mayr, I., Schneider, M., Bode, W., Huber, R., Ries, A. & Kühn, K. (1990). Crystals of the NC1 domain of type IV collagen. *J. Mol. Biol.* **211**, 683–684.
- Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Cryst.* **D55**, 849–861.
- Terwilliger, T. C. & Eisenberg, D. (1983). Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. *Acta Cryst.* **A39**, 813–817.
- Terwilliger, T. C. & Eisenberg, D. (1987). Isomorphous replacement: effects of errors on the phase probability distribution. *Acta Cryst.* **A43**, 6–13.
- Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Cryst.* **A43**, 1–5.
- Tong, L. & Rossmann, M. G. (1993). Patterson-map interpretation with noncrystallographic symmetry. *J. Appl. Cryst.* **26**, 15–21.
- Vagin, A. & Teplyakov, A. (1998). A translation-function approach for heavy-atom location in macromolecular crystallography. *Acta Cryst.* **D54**, 400–402.
- Wilson, A. J. C. (1949). The probability distribution of X-ray intensities. *Acta Cryst.* **2**, 318–321.